

**IMPACT OF POPULATION STRUCTURE ON GENETIC DIVERSITY OF A
POTENTIAL VACCINE TARGET IN THE CANINE HOOKWORM (*ANCYLOSTOMA
CANINUM*)**

Jennifer M. Moser^{*}, Ignazio Carbone[†], Prema Arasu[‡], and Greg Gibson[§]

Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695-7614.

e-mail: statgen.ncsu.edu/ggibson

Key words: *Ancylostoma* secreted protein, K_{ST} , mitochondria, gene flow, test of neutrality

ABSTRACT: *Ancylostoma caninum* is a globally distributed canine parasitic nematode. To test whether positive selection and/or population structure affect genetic variation at the candidate vaccine target, *Ancylostoma* secreted protein 1 (*asp-1*), we have quantified the genetic variation in *A. caninum* at *asp-1* and a mitochondrial gene, cytochrome oxidase subunit 1 (*cox-1*) using the statistical population analysis tools found in the SNAP workbench. The mitochondrial gene *cox-1* exhibits moderate diversity within two North American samples, comparable to the level of variation observed in other parasitic nematodes. The protein coding portion for the C-terminal half of *asp-1* shows similar levels of genetic variation in a Wake County North Carolina sample as *cox-1*. Standard tests of neutrality provide little formal evidence for selection acting on this locus, but haplotype networks for two of the exon regions have significantly different topologies, consistent with different evolutionary forces shaping variation at either end of a 1.3 kilobase stretch of sequence. Evidence for gene flow both among geographically distinct samples suggests that the mobility of hosts of *A. caninum* is an important contributing factor to the population structure of the parasite.

The canine hookworm, *Ancylostoma caninum* is a useful model for the widespread human hookworms *Necator americanus* and *Ancylostoma duodenale*, but can cause severe pathologies in its own right. Heavy infestations in puppies can lead to anemia and death. Conventional treatments to reduce infections must often be repeated, and there is a growing desire to develop vaccines that will target specific genes and gene products of parasites and interrupt the infection cycle. In order to be effective, these reagents must recognize all of the variation observed in the target loci in natural populations. A gene that has been investigated for its potential usefulness as a target of such a treatment is Ancylostoma Secreted Protein-1, which encodes a protein that is secreted into the host during the infection process by the L3 (Hawdon et al., 1996). The *asp-1* gene encodes a two-domain, 431 amino acid protein (Fig. 1A), and is one of several secreted by infective larvae when exposed to host factors and in anticipation of infection (Moser et al., 2005). Since hookworms have evolved to effectively avoid detection by host immune systems, there may be selection pressure to sustain high levels of variation in some or all of the secreted/excreted proteins in hookworm populations (Vacher et al., 2005). This hypothesis can be tested by population genetic surveys of nucleotide diversity, as positive or balancing selection should leave signatures of departure from neutrality (Kreitman, 2000). Furthermore, identification of specific haplotypes that are unique to local populations may have implications for implementation of immunotherapy, specifically vaccines.

Evidence from studies of diverse parasitic animal nematodes suggests that there are several types of population structure in these species (Blouin et al., 1995; Blouin et al., 1999; Qiang et al., 2000; Picard et al., 2004; Nieberding et al., 2005). Population structure in these nematodes, in terms of genetic diversity and divergence, is influenced by levels of gene flow and by host mobility (Blouin et al., 1995; Blouin et al., 1999; Hawdon et al., 2001; Braisher et al.,

2004). Hookworms display intermediate levels of gene flow and structure compared to the highly diverse trichostrongylids and the less diverse, highly structured plant and insect nematodes.

One study comparing the diversity of just two sequences of *asp-1* isolated from two geographic areas showed 97% DNA sequence identity between a Chinese and a Baltimore Maryland strain of *A. caninum* in the coding regions of *asp-1* (Qiang et al., 2000). The study suggested that a reduced level of polymorphism in the second of the two domains in the protein provides evidence for selection against amino acid substitutions there. We elected to focus on this region of the gene in a broader survey of nucleotide diversity that includes both the intron and coding regions of the gene. A mitochondrial gene is often employed as a neutral baseline for comparison of genetic variation and population structure in a species. The mitochondrial *cytochrome oxidase I* locus (*cox-1*) has served as such in previous studies (Tarrant et al., 1992; Anderson et al., 1998; Blouin, 1998; Hawdon et al., 2001; Hu, Chilton et al., 2002; Hu, Hoglund et al., 2002; Hu et al., 2003).

Here, we describe the level of population structure and diversity at *asp-1* and *cox-1* within a local sample of *A. caninum* from Wake County, North Carolina, and a ‘Maryland’ laboratory strain initially isolated in that state. The genomic DNA sequences are analyzed with a comprehensive software package (SNAP workbench) and are compared with a previously published population survey of *cox-1* from Townsville, Australia (Hu, Chilton et al., 2002), and with the reported *asp-1* cDNA sequences from a Chinese *A. caninum* strain (Qiang et al., 2000).

MATERIALS AND METHODS

DNA extractions of whole adult male and female *A. caninum* were carried out by homogenizing individual worms in a digestion buffer consisting of 100 mM NaCl; 10 mM Tris

HCl, pH 8.0; 25 mM EDTA, pH 8.0; 0.5% SDS; 0.1 mg/ml proteinase K. A sample of 118 worms from a Wake County, North Carolina animal shelter (near Raleigh) was collected by P.A., from which we obtained 54 useable sequences. A Maryland sample was maintained in laboratory dogs at the University of Pennsylvania, and was originally collected in the 1970s (kindly provided by Dr. Thomas Nolan, University of Pennsylvania). We used 12 individual worms from this sample for sequencing. Nematode homogenates were prepared as [described by](#) Deleted: described by incubating nematodes at 50 C for 4 hr to dissolve the proteinaceous and cuticular portions of the hookworms (Hu, Chilton et al, 2002). DNA was extracted by standard phenol-chloroform extraction and ethanol precipitation, and stored in 10-20 µl dH₂O at -20 C. Australian samples reported were collected from dogs in Townsville, Queensland, Australia (Hu, Chilton et al., 2002) and the sequences were taken directly from that reference. The GenBank haplotype accession numbers are [AJ407941](#), [AJ407961](#), [AJ407962](#), [AJ407963](#), [AJ407964](#), [AJ407965](#), and [AJ407966](#).

All amplification reactions were performed in a volume of 30 µl in an MJ Research thermocycler using Promega *Taq* polymerase. PCR and sequencing primers for a 373 bp fragment of *cox-1*, were reported in Hu, Chilton et al. (2002) (*cox-1* F: 5'-TTT TTT GGG CAT CCT GAG GTT TAT-3' R: 5'-TAA AGA AAG AAC ATA ATG AAA ATG-3'). The PCR conditions were 95 C 5 min, followed by 30 cycles of 95 C 30 sec, 45 C 30 sec, 72 C 30 sec, with a final extension at 72 C 10 min. Amplified DNA bands were extracted from 0.8% agarose gels using Qiagen gel extraction kits following manufacturer's specifications, and sequenced using BigDye (Applied Biosystems) termination on ABI 3700 sequencers at the NCSU Genome Research Lab. Primer sets were created to amplify three regions of *asp-1* as diagrammed in Figure 1A as follows: *asp-1E* F: 5'-AGG CTT CAC AAC TCT GGT GG-3' R: 5'-ACT GAT

TAA GGA GCA CTG CA-3'; *asp-ID* F: 5'-ATG CAC CAA AAG CAG CTA AA-3' R: 5'-TCC ATA AAG CAG TGG TAA GG-3'; *asp-IC* F: 5'-ACT CCG AAA CGA AAC T-3' R: 5'-TCA TAC ACC ATC TGG AT-3'. For the *asp-IE* region, the PCR conditions were 95 C 5min, followed by 30 cycles of 95 C 30sec, 50 C 30sec, 72 C 30sec, and a 72 C 5min for final extension. The PCR conditions for *asp-IC* and *asp-ID* were 94 C 4min, followed by 36 cycles of 94 C 30sec, 50 C 30 sec, 72 C 2min, with a final extension at 72 C for 5min.

Due to high levels of indel polymorphism in the *asp-I* genomic region, it was not possible to sequence amplified DNA directly, so PCR products for this gene were cloned into the pGEM-T vectors using the Promega PCR cloning kit as per manufacturer's instruction. Plasmids were isolated from bacteria with Qiagen miniprep kits, restriction digests performed to confirm the presence of the appropriate length insert. Inserts were sequenced with the vector specific T7 and SP6 primers. The number of sequences for regions *asp-IE*, *asp-ID*, and *asp-IC* were 70, 8, and 22, respectively.

Sequence analysis was performed using the ContigExpress module of Vector NTI (InforMax, MD) to manually edit the sequence ABI files. To determine the coding and non-coding portions of the *asp-I* regions, BLASTn (Altschul et al., 1990) was used to compare genomic sequence to the known coding regions of the *asp-I* mRNA represented by GenBank accession ID: AF13229 (Fig. 1B). This figure includes the cDNA sequence along with the shortest sequenced introns. SNAP Workbench (Price and Carbone, 2005; Aylor et al., 2006), a platform combining numerous population genetic and phylogenetic software programs, was used to create multiple sequence alignments (CLUSTALW; Thompson et al., 1994) and allowed for transformation of sequence data into different formats, including NEXUS (Maddison et al., 1997) and PHYLIP (Felsenstein, 2004). Haplotype networks were created with GT Miner

version 1.16, a component of SNAP workbench (D. Brown and I. Carbone, pers. comm.) based on a maximum parsimony approach (Cassens et al., 2005). Most parsimonious trees were inferred with the Discrete Character Parsimony Algorithm, version 3.6b implemented in PHYLIP (Felsenstein 2004). SNAP was also used to test for population subdivision and migration using Hudson's tests (Hudson et al., 1992) and MDIV (Nielsen and Wakeley, 2001), respectively. MDIV tests for gene flow over the entire coalescent history of a pair of populations and therefore captures both historical and ongoing migration between the two populations. Specifically, MDIV applies coalescence simulation methods to estimate divergence time and integrated migration rates from population 1 to population 2 and migration from population 2 to population 1. Posterior likelihood plots of divergence time and migration were plotted using gnuplot (<ftp://ftp.gnuplot.info/pub/gnuplot>). DnaSP v.4.0 (Rozas et al., 2003) was used to estimate population parameters representing nucleotide diversity, haplotype diversity, and Watterson's estimation of theta. It was also used to perform Tajima's, Fu and Li's, and Fu's tests of neutrality (Tajima, 1989; Fu and Li, 1993; Fu, 1997). MEGA3 (Molecular Evolutionary Genetics Analysis; Kumar et al., 2004) was used to infer gene genealogies. We tested for linkage disequilibrium (LD) among polymorphisms in *asp-1C* and *asp-1E* using significance tests implemented in TASSEL (<http://www.maizegenetics.net>). To address potential errors from amplification and sequencing, the *asp-1* regions were reanalyzed after further collapsing the haplotype sequences that differed by fewer than 2 mutations.

RESULTS

Cytochrome oxidase subunit-1

The mitochondrial locus *cox-1* is traditionally used in population genetics studies because it exhibits absence of recombination and a relatively high mutation rate. Table I summarizes the

nucleotide diversity statistics for the Wake County (WC) and Maryland (MD) samples and compares these with the study of Australian samples. The estimate of nucleotide diversity (π), which represents the proportion of pairwise nucleotide differences among sequences, is 0.014 in the Wake County sample, 0.001 in the Maryland sample, and 0.055 in the Australian sample. Since the Maryland population was expanded from a small founder group, there was likely a loss of genetic variation resulting in the relatively low π (Gasnier and Cabaret, 1998). The Australian sample has a much higher π than either of the American samples, but this is mostly attributable to 3 haplotypes that differ by at least 22 nucleotides (out of 373) from the other 4 Australian haplotypes. Diversity among the remaining four Australian haplotypes is similar to that observed in the North American samples.

The 62 sequenced *cox-1* amplicons and 38 Australian sequences represent 19 haplotypes as shown in Figure 2. Phylogenetic analysis in the SNAP workbench results in 56 equally most parsimonious trees possible from the 19 haplotypes. Most of the trees contain 3 major clusters clearly visible in the haplotype network (Fig. 3). The 3 highly divergent Australian haplotypes (H3, H4 and H5) are visible at the top of the network, while 4 less divergent Wake County haplotypes (H2, H6, H7, and H14, each separated by at least 5 nucleotides from the other haplotypes) group at the bottom of the figure. The majority of the sequences form a network with members from each population sample, in which each haplotype is just one or two nucleotides away from its nearest neighbor. The 3 most common haplotypes are H9 with 25 sequences, H7 with 14 and H1 with 8, the latter two haplotypes consisting entirely of Wake County samples.

A version of Hudson's test (Hudson et al., 1992) implemented in the SNAP workbench was used to test for population subdivision between samples. K_{ST} values, the statistic of

Hudson's test, are based on Wright's fixation index and indicate differentiation among samples of a species. At the *cox-1* locus, the K_{ST} value between Maryland and Wake County samples is low but significant 0.048 ($P = 0.04$), suggesting low levels of genetic differentiation and moderate levels of gene flow (Table II). The results from MDIV, a program from the SNAP workbench which tests whether both populations descended from a panmictic population and were then either connected by migration or completely isolated also support the idea of gene flow between the two populations. Figure 4A shows the posterior distribution of the point parameter M reaches a maximum at 3.4 for this comparison. Combining the sequences from Maryland and Wake County hookworms into a mid-Atlantic (U.S.A.) sample facilitates comparison with the Australian sample. Comparing U.S.A. to Australian samples results in a larger K_{ST} value (0.177, $P < 0.001$), indicating a very high level of genetic differentiation between the 2 populations, consistent with a low level of gene flow between them (Table II). There is supporting evidence for gene flow between the U.S.A. and Australian samples, as seen from the MDIV data, where the maximum value of M along the distribution is 0.60 (Fig. 4B); a maximum M of greater than 0 suggests migration, whereas $M=0$ supports the hypothesis of isolation.. This inference is strengthened if the 3 highly divergent Australian haplotypes are removed from the analysis, in which case the 2 populations are as similar as Wake County and Maryland. The data suggest that Australian and U.S.A. populations descended from a common ancestral population not too distantly in the past ($T = 0.4$, measured in coalescent time units of N_e generations, where N_e is the effective population size), and a low and constant level of gene flow accompanied that population divergence. There is evidence for deviation from mutation-drift equilibrium at *cox-1* in the Wake County sample of *A. caninum*, as Fu and Li's D^* test statistic is significantly negative (Table I), suggesting an excess of low frequency polymorphisms within

the population resulting from background selection (Fu and Li, 1993). This result must be tempered by the recognition that a similar test statistic, Fu and Li's F^* , is not significant.

Ancylostoma Secreted Protein-1

The levels of π across the locus are comparable to the mitochondrial gene *cox-1* when the highly polymorphic intron sequences are included, but considerably lower when only considering the coding sequences. The 3 consecutive regions of *asp-1* highlighted in Figure 1A designated *asp-1C*, *asp-1D*, and *asp-1E* have π values of 0.033, 0.004, and 0.016, respectively (Table III).

The *asp-1E* region spans approximately 400 nucleotides, contains 2 introns and is located at the 3'-end of the gene. Due to editing for proper sequence alignment, the multiple sequence alignments analyzed are only as long as the highest quality, shortest sequence. Seventy *asp-1E* sequences were obtained, 57 from Wake County, and 13 from Maryland, and these were contrasted with the existing Chinese sequence (GenBank accession [AAD31839](#)). A strict consensus tree of the *asp-1E* coding regions shows a comb-like structure that is the hallmark of a recent population expansion or an unresolved phylogeny that is attributable to the small number of segregating sites in the coding regions (data not shown). The Maryland samples do not form a unique cluster, but instead are a part of the unresolved portion of the tree. The network of sixteen haplotypes shown in Figure 5A includes a large interior haplotype (H12; 30 sequences), which contains both Maryland and Wake County sequences with several haplotypes radiating from it. The Chinese haplotype, H2, the most divergent haplotype in the network, is 3 nucleotide substitutions away from a non-sampled inferred interior node and 4 mutational steps from H12.

Hudson's test was performed using the SNAP workbench to assess population subdivision among Wake County and Maryland *A. caninum* at *asp-1E*. The K_{ST} value of 0.016 is

low but significant ($P=0.04$) (Table II), suggesting little genetic differentiation and moderate levels of gene flow among the population samples. The results from MDIV analysis implemented in SNAP Workbench also support the finding of moderate gene flow ($M=1.74$) (Fig. 4C). This is consistent with the results from the *cox-1* locus.

The 715 nucleotide *asp-1C* region contains 3 exons and 3 introns, with the latter being highly polymorphic for both indels and single nucleotide polymorphisms. In the Wake County sample, the overall π is 0.033, which is twice as high as the *asp-1E* and *cox-1* loci, and 22 sequences are represented by 20 haplotypes. Despite the large number of segregating sites, there is only 1 most parsimonious tree with all interior branches showing $>98\%$ bootstrap support (Fig. 6). There are 3 main clusters of haplotypes, with a region of unresolved phylogeny in the center cluster, containing 12 sequences in 10 haplotypes. The haplotype network in Figure 5B displays this region of polytomy surrounding an interior node haplotype (H18) consisting of 3 sequences. Tests of neutrality summarized in Table III did not suggest any deviation from neutrality in this region. Independent analysis of genomic sequences isolated from single worms indicated that there is very little linkage disequilibrium between the *asp-1C* and *asp-1E* regions (data not shown).

The internal *asp-1D* region was difficult to sequence, likely because of the high polymorphism content. Only 8 high quality sequences were available, all from Wake County. This small sample contains only a few segregating sites and correspondingly exhibits a low level of nucleotide diversity ($\pi=0.004$). Because of the low π and small number of haplotypes ($h=3$) neither a strict consensus tree nor a haplotype network was inferred for this region. Further collapsing haplotypes in *asp-1* networks by removing sequences that differ by fewer than two

nucleotides from other sequences yielded no substantial differences from when all sequences are considered, and so the potential PCR and sequencing errors are assumed to be minimal.

DISCUSSION

Prior to this study, population genetic analysis of the canine hookworm was restricted to a small survey of nucleotide diversity in the mitochondrial gene *cox-1* and comparison of just 2 cDNA sequences of the *asp-1* gene. Our survey of 2 mid-Atlantic samples has placed these studies in a broader perspective, and suggests that for the most part genetic variation in the species is globally distributed. Intriguingly, there are 2 divergent groups of *cox-1* haplotypes in the Australian and Wake County samples, mainly due to the relatively large number of nucleotide substitutions among the Australian samples, but most of the sequences share a recent common ancestor in a central network. At *asp-1E*, the single Chinese sample is the most divergent sequence, but it is not significantly different given the diversity observed in Wake County and Maryland. The nuclear and mitochondrial analyses are in agreement suggesting that there has been migration between the populations, and that there is little evidence for strong population structure.

In mitochondrial *cox-1*, the average nucleotide diversities range from a low of 0.001 in Maryland worms to a high of 0.055 in the Australian samples. It has been suggested the divergence within the Australian hookworms relates to sampling of worms with different infectivity strategies (Hu, Chilton et al., 2002). At any given nucleotide site in *cox-1*, pairs of alleles from the Maryland sample will match more than 99.9% of the time. This percentage decreases to 98.6% for Wake County worms, and to an even lower 94.5% in the Australian *A. caninum*. The Maryland sample shows lower diversity because of its confinement in a laboratory setting that has enforced inbreeding and prevented the gene flow that occurs in a

“wild” population. The levels of nucleotide diversity for Wake County samples are on par with other studies of mtDNA diversity in parasites with vertebrate hosts (Blouin et al., 1995, Blouin et al., 1999, Hawdon et al., 2001, Brashier et al., 2004). Despite the existence of at least three clusters of haplotypes in *cox-1* of *A. caninum*, the nucleotide diversity is mostly consistent with a neutral model of evolution.

The other gene of interest, the nuclear *asp-1*, is a highly variable gene. Its average nucleotide diversity across the analyzed regions is 0.020, and the ranges of π in three segments of the gene were 0.004 to 0.033, represent an almost 10-fold difference. Antigenic diversity may exist, though it is not known whether this may impact the development of vaccines targeted against the ASP-1 protein. There is no a priori reason why amino acid sites that are conserved because of their contribution to function of the secreted protein should also be the primary sites that promote antigenicity of the protein. Given that there is amino acid diversity, it does become relevant to determine which, if either, domain is more functional in conferring resistance to an *A. caninum* infection and/or response to vaccination. There appears to be a nominal level of variation in *asp-1* coding regions between Shanghai and Baltimore samples of *A. caninum*, with most of the variation occurring in domain I (Qiang et al., 2000). Our analysis detected higher levels of variation in the coding portions of Wake County and Maryland DNA than was previously found. There is considerably higher nucleotide diversity in the introns, implying, not surprisingly, that negative selection constrains polymorphism in the exons.

An unexpected finding concerns the independence of haplotype network structure within short segments of the *asp-1* locus. The 2 networks in Figure 5 have very different topologies, i.e., *asp-1E* is a very typical topology for a region with low diversity, while *asp-1C* shows an unusual bipartite topology. One cluster of haplotypes radiates around the most frequent

haplotype, while another 10 haplotypes spread out as a long tail, with several nucleotides in each haplotype different from the others. Standard tests of neutrality do not imply that selection has shaped this variation, and it is possible that the major cluster on the left side of the network represents a fairly recent expansion. It suggests that more extensive sampling may uncover local population structure such as that described for *N. americanus* in China (Hu et al., 2003). A further implication of these results is that there is very little linkage disequilibrium in the *A. caninum* genome. This inference is confirmed by limited bi-directional sequencing of genomic *asp-1* products spanning the locus, as 4 gamete tests imply that recombination breaks up associations between common polymorphisms separated by just 2 kilobases (data not shown). In this respect, the canine hookworm is more like *Drosophila melanogaster* than humans or the soil nematode *Caenorhabditis elegans*, both of which show extensive linkage disequilibrium over tens of kilobases (Reich et al., 2001; Gabriel et al., 2002; Cutter, 2005). A further consequence is that association mapping of genotypes that correlate with infectivity or antigenicity must be focused on individual SNPs rather than tagging of haplotypes. A corollary is that if positive selection is shaping variation in the *asp-1* locus, its impact on diversity has been restricted to just a few hundred nucleotides surrounding any focal site, and must therefore necessarily be relatively weak. Nevertheless, the relatively high levels of polymorphism and absence of linkage disequilibrium are excellent conditions favoring the use of population genetic strategies to identify regions of the genome that shape the parasite's response to antihelminthics and immunotherapeutic options like vaccination.

ACKNOWLEDGEMENTS

J.M. was supported by an IGERT Training Grant of the NSF, and the research was facilitated by NIH grant 1-R01-GM61600 to GG.

LITERATURE CITED

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**: 403-410.
- Anderson, T. J., M. S. Blouin, and R. N. Beech. 1998. Population biology of parasitic nematodes: applications of genetic markers. *Advances in Parasitology* **41**:219-283.
- Aylor, D.L., E. Price, and I. Carbone. 2006. SNAP: Combine and Map modules for multilocus population genetic analysis. *Bioinformatics* **22**: 1399-1401.
- Blouin, M. S. 1998. Mitochondrial DNA diversity in nematodes. *Journal of Helminthology* **72**: 285-289.
- _____, J. Liu, and R. E. Berry. 1999. Life cycle variation and the genetic structure of nematode populations. *Heredity* **83** (Pt 3): 253-259.
- _____, C. A. Yowell, C. H. Courtney, and J. B. Dame. 1995. Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* **141**: 1007-1014.
- Braisher, T. L., N. J. Gemmell, B. T. Grenfell, and W. Amos. 2004. Host isolation and patterns of genetic variability in three populations of *Teladorsagia* from sheep. *International Journal for Parasitology* **34**:1197-1204.
- Cassens, I., P. Mardulyn, and M. C. Milinkovitch. 2005. Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? *Systems Biology* **54**:363-372.
- Cutter, A. D. 2006. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. *Genetics* **172**(1): 171-184.
- Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6 in D. b. t., ed, Department of Genomics Sciences, University of Washington, Seattle, Washington.

Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915-925.

_____, and W. H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**: 693-709.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, and D. Altshuler. 2002. The structure of haplotype blocks in the human genome. *Science* **296**: 2225-2229.

Gasnier, N. and J. Cabaret. 1998. Stable and unstable environments influence the genetic diversity of the nematode *Teladorsagia circumcincta*, a parasite of small ruminants. *Parasitology Research* **84**: 676-681.

Hawdon, J. M., B. F. Jones, D. R. Hoffman, and P. J. Hotez. 1996. Cloning and characterization of Ancylostoma-Secreted Protein. *Journal of Biological Chemistry* **271**: 6672-6678.

_____, T. Li, B. Zhan, and M. S. Blouin. 2001. Genetic structure of populations of the human hookworm, *Necator americanus*, in China. *Molecular Ecology* **10**: 1433-1437.

Hu, M., N. B. Chilton, Y. G. Abs El-Osta, and R. B. Gasser. 2003. Comparative analysis of mitochondrial genome data for *Necator americanus* from two endemic regions reveals substantial genetic variation. *International Journal for Parasitology* **33**: 955-963.

_____, _____, X. Zhu, and R. B. Gasser. 2002. Single-strand conformation polymorphism-based analysis of mitochondrial cytochrome c oxidase subunit I reveals significant substructuring in hookworm populations. *Electrophoresis* **23**: 27-34.

- _____, J. Hoglund, N. B. Chilton, X. Zhu, and R. B. Gasser. 2002. Mutation scanning analysis of mitochondrial cytochrome c oxidase subunit 1 reveals limited gene flow among bovine lungworm subpopulations in Sweden. *Electrophoresis* **23**: 3357-3363.
- Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583-589.
- Kreitman, M. 2000. Methods to detect selection in populations with applications to the human. *Annual Review of Genomics and Human Genetics* **1**: 539-559.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinformatics* **5**: 150-163.
- Maddison, D. R., D. L. Swofford, and W. P. Maddison. 1997. NEXUS: an extensible file format for systematic information. *Systems Biology* **46**: 590-621.
- Moser, J. M., T. Freitas, P. Arasu, and G. Gibson. 2005. Gene expression profiles associated with the transition to parasitism in *Ancylostoma caninum* larvae. *Molecular and Biochemical Parasitology* **143**: 39-48.
- Nieberding, C., R. Libois, C.J. Douady, S. Morand, and J. R. Michaux. 2005. Phylogeography of a nematode (*Heligmosomoides polygyrus*) in the western Palearctic region: persistence of northern cryptic populations during ice ages? *Molecular Ecology* **14**: 765-779.
- Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**: 885-896.
- Picard, D., O. Plantard, M. Scurrah, and D. Mugniery. 2004. Inbreeding and population structure of the potato cyst nematode (*Globodera pallida*) in its native area (Peru). *Molecular Ecology* **13**: 2899-2908.

- Price, E. W., and I. Carbone. 2005. SNAP: workbench management tool for evolutionary population genetic analysis. *Bioinformatics* **21**:402-404.
- Qiang, S., Z. Bin, X. Shu-hua, F. Zheng, P. Hotez, and J. M. Hawdon. 2000. Variation between ASP-1 molecules from *Ancylostoma caninum* in China and the United States. *Journal of Parasitology* **86**:181-185.
- Reich, D. E., M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti, D. J. Richter, T. Lavery, R. Kouyoumjian, S. F. Farhadian, R. Ward, and E. S. Lander. 2001. Linkage disequilibrium in the human genome. *Nature* **411**: 199-204.
- Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496-2497.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA Polymorphism. *Genetics* **123**: 585-595.
- Tarrant, C. A., M. S. Blouin, C. A. Yowell, and J. B. Dame. 1992. Suitability of mitochondrial DNA for assaying interindividual genetic variation in small helminths. *Journal of Parasitology* **78**: 374-378.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.
- Vacher, C., S. P. Brown, and M.E. Hochberg. 2005. Avoid, attack or do both? Behavioral and physiological adaptations in natural enemies faced with novel hosts. *BMC Evolutionary Biology* **5**: 60.

FIGURE 1. (A) Schematic of sequenced regions of the *asp-I* gene. Three distinct regions, labeled *asp-IC*, *asp-ID*, and *asp-IE* are mostly contained within Domain 2 and were sequenced using designed primers in materials and methods section. The rectangles connected by lines represent exons and introns, respectively. The dashed line represents the 5' Domain 1 of *asp-I*, which was not sequenced. The bold dashed line connecting *asp-ID* and *asp-IE* represents an unsequenced region. (B) *asp-I* DNA sequence. Nucleotides sequenced in the study are bolded. The 7 introns were identified by aligning sequenced genomic DNA with cDNA sequence (GenBank accession: AF13229) and are underlined. The introns included represent the shortest possible intron sequences observed in the data.

FIGURE 2. *cox-I* haplotype SNP matrix. Sixty-two *A. caninum* individuals from Wake County and Maryland were sequenced at the *cox-I* locus and the resultant sequences, along with 38 samples representing 7 haplotypes (Hu, Chilton et al., 2002) were collapsed into 19 new haplotypes using the SNAP Workbench. Indels and infinite sites violations were removed from the analysis. The matrix shows the position of single nucleotide polymorphisms (SNPs) in the multiple sequence alignment, e.g., the SNP at site 1 is located at position 16 in the alignment, the identity of the SNPs relative to the consensus sequence, and displays whether the nucleotide changes are transversions (v) or transitions (t). The number of individual sequences contained in each haplotype is noted in parentheses.

FIGURE 3. Unrooted *cox-I* haplotype network. In this network, generated by the GT Miner component of the SNAP Workbench, the 19 haplotypes from the *cox-I* analysis (calculated by removing both indels and infinite-sites violations) are shown as oval nodes connected by

branches whose length correspond to the number of nucleotide differences between them. The black nodes represent the Australian haplotypes, the unshaded nodes are Maryland and Wake County haplotypes, and the gray nodes represent unsampled haplotypes. H9 is the largest haplotype, and contains 25 *A. caninum* samples, including all 13 of the Maryland samples.

FIGURE 4. MDIV results from the SNAP Workbench. The graphs show the posterior probability distribution of migration and divergence time between each of 2 geographic samples for a particular gene region. Ten independent runs were simulated under the assumption of an infinite sites model, with 2,000,000 steps in the chain and a burn-in time of 500,000 using a different starting random number seed for each run. The distributions for migration and time of divergence were similar between runs. The data were plotted with Gnuplot version 4.6. (A) posterior probability distribution for migration (M) and divergence time (T) between Wake County and Maryland samples at the *cox-1* locus; (B) between US and Australia samples at *cox-1*; (C) between Wake County and Maryland samples at *asp-1E*

FIGURE 5. *asp-1* haplotype networks. Networks were generated using the GT Miner component of SNAP Workbench to display the mutational connections between haplotypes. The haplotypes are represented by oval nodes, connected by branches that indicate the number of nucleotide differences between them. Empty nodes represent unsampled haplotypes that are inferred to exist, or have existed at some time, but have not been sampled. A) *asp-1E* coding region haplotype network. H1 contains both Maryland and Wake County samples. H2 represents the Chinese *asp-1* (GenBank accession [AAD31839](#)); B) *asp-1C* haplotype network contains only Wake County samples.

FIGURE 6. Maximum parsimony tree of *asp-IC* haplotypes inferred using PHYLIP in the SNAP Workbench and displayed with MEGA3. The 20 distinct haplotypes were inferred by collapsing 22 alleles excluding infinite sites violations.

*Current address: Laboratory of Allergic Diseases, National Institutes of Health, Bethesda, Maryland 20892.

†Center for Integrated Fungal Research, Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina 27695-7244.

‡Department of Molecular Biomedical Sciences, North Carolina State University College of Veterinary Medicine, Raleigh, North Carolina 27695-8401.

§To whom correspondence should be addressed.

Figure 1A. The gene structure of *asp-I*.

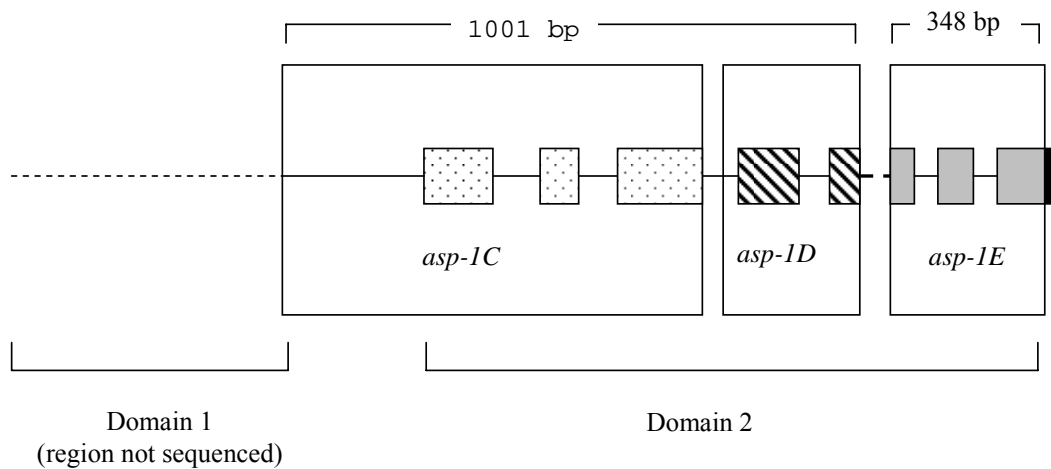


Figure 1B. *asp-1* sequence.

```

atgtttttctc ctgtagtcgt cagtgtgatt ttcacaatcg ccatctgcga tgcgtctcca
gcaagagaca gcttcggctg ctctaacagt gggataactg acaaggaccg gcaagcattc
ctcgacttcc acaacaatgc ccgtcgacgg gttgcgaaag gccttgagga tagcaactcc
ggcaactga atccagcgaa gaacatgtac aagctgtcat gggactgtgc aatggaacag
cagcttcagg atgccatcca gtcatgcccc agcggctttg ctgggattca aggtgttgca
cagaatacaa tgagctggtc aagctctggt ggattccccg atccatcggg aaagatagaa
ccaacgctct ccggtcgggt gagtgggtgc aaaaagaacg gtgtcggccc ggacaacaaa
tacaacgggt gtggtctctt cgccctctct aacatggtat actocgaaac gacgaaactt
gggtgcgcct acaaggtttg cggcactaaa ctggcggttt cgtgcatcta taatggagtc

```

Domain 1

```

tgcaatgaga aatgcaaac aggataattt tcacatgttg actttcacct caaaaccaat
aggaatagc cccgcattgc agaaagtgcg gaaaaaggca caaacaggat tatcaatgac
aactattgcg aataaaatcg tgtaaaattc ttcccttaag aaaccttcta cttaccggta
cgaattattg ggacaaacaa cttcgaaccc ctactctcca ttattacaat acttatttct
agcgggttaca tcacaaatca acctatgtgg gagacaggtc aggcttgcca gacaggagca
gactgctcca cttacaagaa ctcaggctgc gaggatggcc tttgcacgaa gggaccagat
gtgccaggtc tgttcatagc gtattttgcg gccacatcc ttcaaaacttc gcctaaattt
aagataaaaag ctacactctg aaaatatgta tgctttcacc agttcagaaa caaaccagca
gtgccctca aacaccggaa tgactgattc agtcagagat actttcctat cgggtgcacaa
tgagttcagg tatgttcagg ggctcaagtc tttaacggga cttagagggcg tctgaaat
gcctatcgag cttcccttac agatcgagtg ttgcccagg tctggaaccgg acgctctgg
gcggaaatgc accaaaagca gctaaaatgc tcaagataaa aaaactctcaa taagagtgg
cctcatttca ggtgtatgac tgtgaagtgg aagcatcggc catcagacatg gaaataaat
gcgtctatca acattctcat ggtgaagaca gacctggact aggagaaaaca tctacaaga
ctagtgtact caaattcgac aagaacaaa gaggcaagca ggcattgattg cgactacag
tacatcattt catgccctac tacaactgat tccgtcattt aaggcttcaca actctggtg
gaatgagtta aaagagtacg gcgttgccc atccaacgtc cttaccactgc gttatggaa
tagaccaaac atgcagattg gtcactacac ccagggtgaac ttgctcgttcc acgcttttt
ttttcaatcc aaaaataaaa atctgccccg atttagatgg catgggacacc acctacaaa
cttggtatgtg cagttgtttt ctgcaatgat ttcacattcg gcgtttgtcag tatgggcca
gggtacgttc tacaagaagt tacatctaca gataccgttt gctacgattct actagtggc
tttattttgtg ttttgcagag gcaattacat ggggtcatgtc atctacactat gggccagcc
gtgctctcag tgttcgcctg gtgctaattg cagcgtgacc gaagggtgtgca gcgctcctt
aatcagtcaa caataaatat cttacagtg

```

Domain 2

Position		11111111111111222222222222222333333333
		1122356689012233455567888011223456677790223344556
		6814916913240658734627369103154321703676470958470
Site Number		111111111122222222223333333333444444444
		1234567890123456789012345678901234567890123456789
Consensus		GTAGAATATAGGATTGGAGTGATAATAGGAGAGAAGGAGTGAGAATCGT
Site Type		tvttttvtttttttttttttvtttttttvtvtvtttttttttttttttttt
H1	(8)G...A.....C
H2	(2)	A...G.....GC.....G...G.....C...
H3	(8)	.AGA.GAG.GAA...AA..C.G.G.CT.....G..AG.CA..GG.TAC
H4	(1)	.AGAGGAG.GA...AA...G.G..T...G..AG.C.G.GG.TAC
H5	(2)	.AGA.GAG.GA...AA..C.G.G.CT.....G..AG.CA..GG.TAC
H6	(4)	...G...C...GC.....G...G.....C...
H7	(14)C...GC.....G...G.....C...
H8	(1)C.....G.....
H9	(25)C.....A.....
H10	(5)C.....G.....A...G.....
H11	(1)C.....C.....A...A.....
H12	(3)C.....T...
H13	(11)C...G...G.....
H14	(1)C...GC...T.....TG...G.....C...
H15	(1)A.....G..A.....TAC
H16	(1)CC.....T...A...A...A..A..
H17	(1)A.....A.....A.....
H18	(1)A.....
H19	(10)A.....

Figure 3. Unrooted *cox-1* haplotype network.

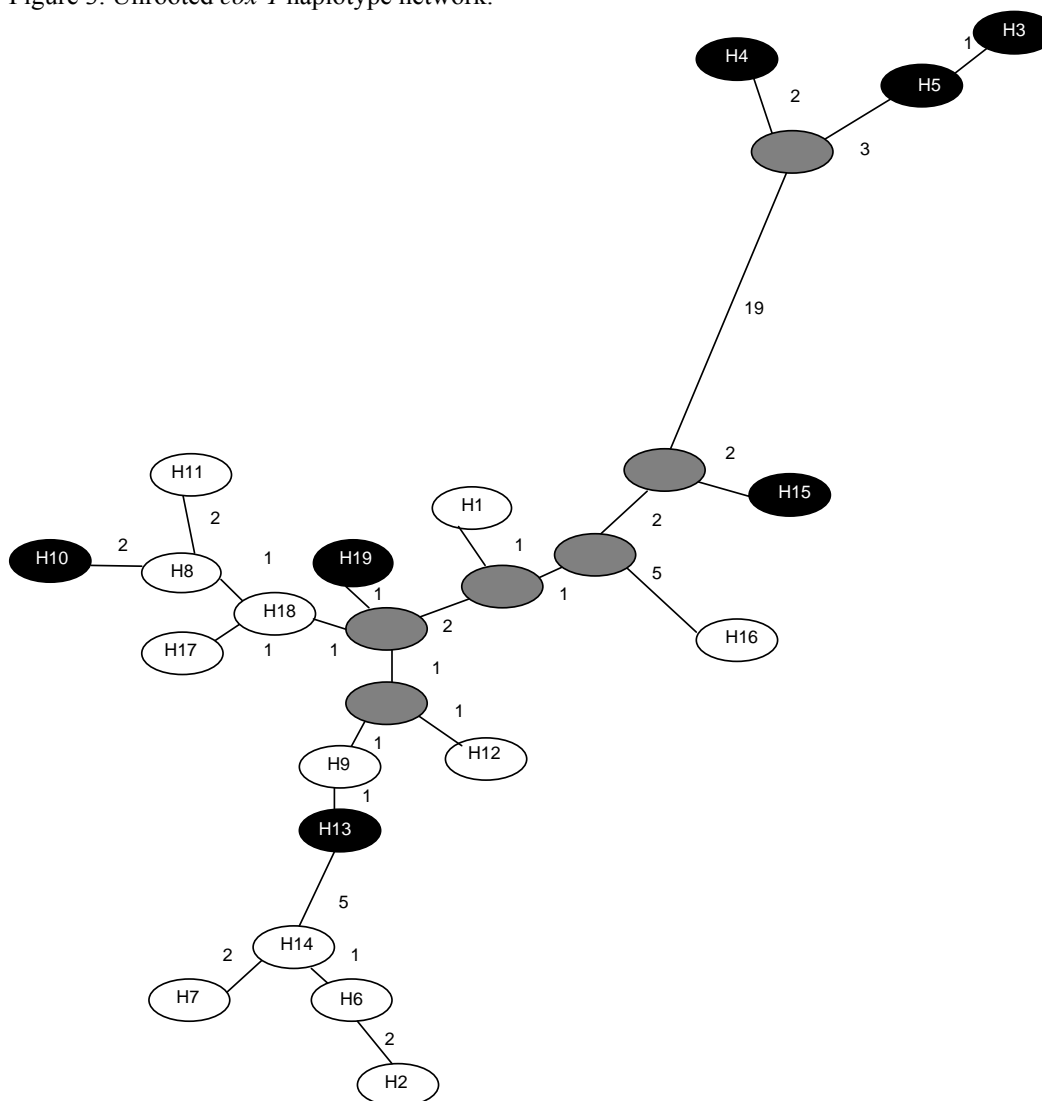
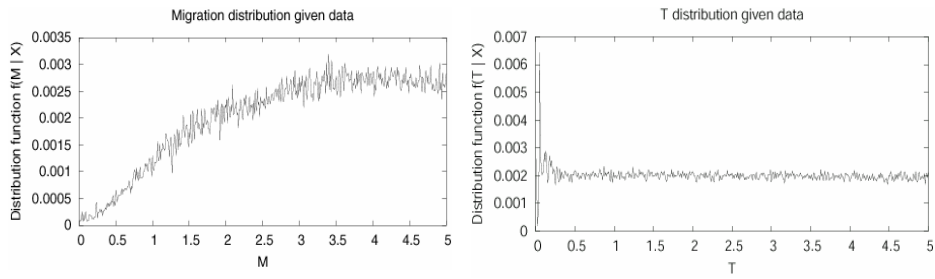
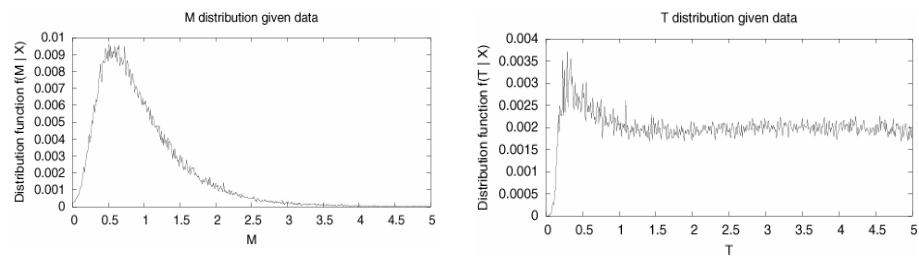


Figure 4. MDIV results for detecting genetic isolation between samples.

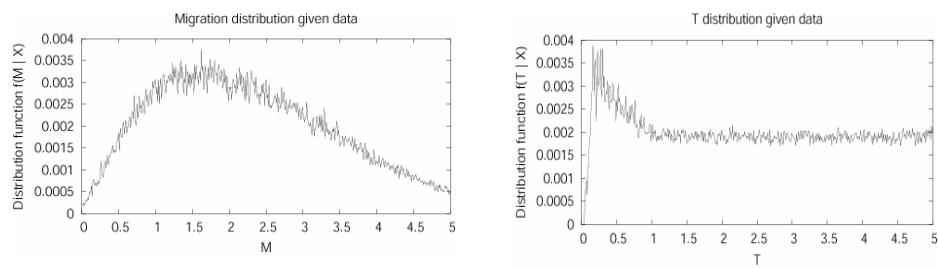
A. Wake County/Maryland *cox-1*



B. U.S.A./Australia *cox-1*



C. Wake County/Maryland *asp-1E*



A. *asp-1E*

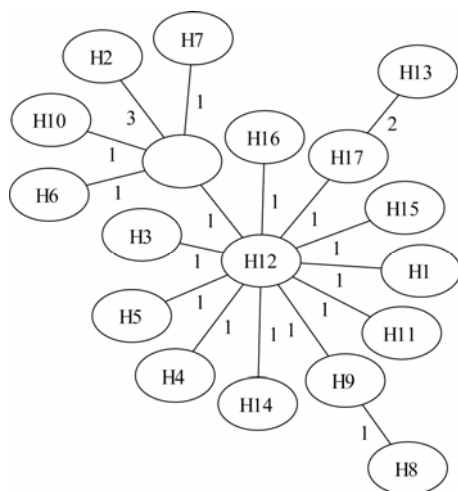


Figure 6. Unrooted *asp-1C* haplotype maximum parsimony tree.

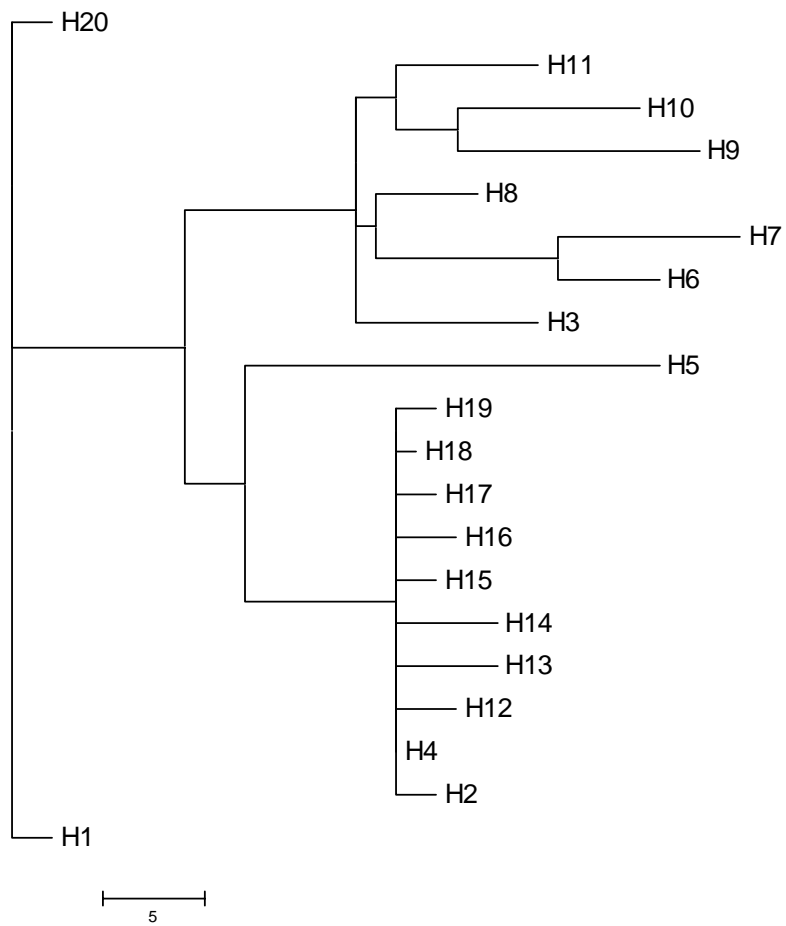


Table I. *cox-I* population parameters.

Sample	Mutations	Sample size	Haplotypes H diversity	π	θ /locus θ /site	Tajima's D	Fu & Li's D^*	Fu & Li's F^*	Fu's F
WC	30	54	14 0.828±0.031	0.014±0.001	6.145 0.017	-0.627ns	-2.556*	-2.218 ns	-0.149
MD	2	8	2 0.250±0.18	0.001±0.001	0.771 0.002	-1.310 ns	-1.410 ns	-1.514 ns	-0.396
Australia	42	38	7 1.0±0.076	0.055±0.008	17.96 0.048	0.778 ns	0.554 ns	0.669 ns	0.762

ns, not significant

Table II. Hudson's test for subdivision and results of MDIV test for isolation.

Samples compared	locus	K_S	K_T	K_{ST}	P -value	Genetic differentiation?	Max M (MDIV)
WC/MD	<i>cox-1</i>	4.175	4.841	0.048	0.04	yes	3.40
U.S.A*/AUS	<i>cox-1</i>	5.480	6.658	0.177	<0.001	yes	0.60
WC/MD	<i>asp-1E</i>	4.340	4.308	0.016	0.04	yes	1.74

*U.S.A. sample is created by combining Wake County (WC) and Maryland (MD) samples

Table III. *asp-I* population parameters.

A. Entire sequence.

	Sites [*]	Mutations	Sample size	Haplotypes H diversity	π	θ /locus θ /site	Tajima's <i>D</i>	Fu and Li's <i>D</i> *	Fu & Li's <i>F</i> *	Fu's <i>F</i>
<i>asp-IE</i> [†]	348	43	70	33 0.932±0.017	0.016±0.001	9.33 0.028	-1.400 ns	-0.976 ns	-1.360 ns	-16.75
<i>asp-ID</i>	286	3	8	3 0.679±0.12	0.004±0.013	1.15 0.004	-0.431 ns	-0.630 ns	-0.642 ns	0.270
<i>asp-IC</i>	715	129	22	20 0.987±0.020	0.033±0.004	35.39 0.043	-1.436 ns	-2.011 ns	-2.147 ns	-3.808

B. Coding regions.

	Sites	Mutations	Sample size	Haplotypes H diversity	π	θ /locus θ /site	Tajima's <i>D</i>	Fu & Li's <i>D</i> *	Fu & Li's <i>F</i> *	Fu's <i>F</i>
<i>asp-IE</i> [†]	208	16	70	16 0.779±0.043	0.007±0.001	3.321 0.016	-1.704 ns	-0.781 ns	-1.317 ns	-10.318
<i>asp-ID</i>	191	0	8	1 0.0	0	0	n/a	n/a	n/a	n/a
<i>asp-IC</i>	289	31	22	18 0.957±0.037	0.014±0.002	8.504 0.029	-1.997*	-2.842*	-3.020*	-12.751

C. Non-coding regions.

	Sites	Mutations	Sample size	Haplotypes H diversity	π	θ /locus θ /site	Tajima's <i>D</i>	Fu & Li's <i>D</i> *	Fu & Li's <i>F</i> *	Fu's <i>F</i>
<i>asp-IE</i> [†]	125 [‡]	26	70	22 0.889±0.024	0.028±0.002	5.412 0.044	-1.704 ns	-1.124 ns	-1.169 ns	-8.551
<i>asp-ID</i>	62	3	8	3 0.679±0.015	0.017±0.006	1.157 0.019	-0.431 ns	-0.632 ns	-0.642 ns	0.268
<i>asp-IC</i>	425	97	22	19 0.974±0.028	0.046±0.006	26.61 0.065	-1.626 ns	-1.751 ns	-3.020 ns	-3.399

*Includes gaps and missing data.

[†]Contains alleles from both Wake County and Maryland *A. caninum*.

[‡]Sequence alignment gaps removed.

n/a, no sequence variation